

Introdução

No âmbito da crescente influência das redes sociais na formação de opinião e na disseminação de conteúdo, o projeto em questão almeja aprofundar-se na dinâmica específica das interações relacionadas à culinária no Facebook. Com o intuito de decifrar os intrincados padrões dessa vasta rede, concentramo-nos na detecção de influenciadores em páginas verificadas dedicadas ao universo gastronômico.

Apresentação da Base de Dados

A base de dados utilizada é da categoria Sparse Networks disponível na coleção Social Networks, sendo os dados referentes ao mês de novembro de 2017. A representação da rede é feita através de vértices que correspondem às páginas e arestas que simbolizam as interações de "likes" mútuos entre as páginas. A base de dados contém 620 vértices e 2100 arestas, com um grau médio da rede estimado em 6.78. O grafo é não-direcionado, e os vértices são caracterizados pelos atributos de Id da página, nome da página no Facebook e um novo Id para o vértice, sendo esse utilizado no arquivo de arestas. A obtenção dos dados foi realizada através do Network Data Repository, uma plataforma interativa para dados de redes: <https://networkrepository.com/>

Limitação do Escopo e Definição do Problema

O problema a ser abordado é a detecção de influenciadores na rede social Facebook, considerando páginas verificadas relacionadas à culinária. O objetivo é identificar atores-chave que exercem impacto significativo no compartilhamento e disseminação de conteúdo culinário, utilizando o grafo representativo da rede.

Entrada Esperada:

- Grafo representativo da rede social Facebook, com vértices correspondendo a páginas e arestas representando interações de "likes" mútuos entre as páginas. Um arquivo representando os vértices, e outro as arestas.
- Atributos dos vértices, incluindo ID da página, nome da página no Facebook e um novo ID para o vértice.

Saída Esperada:

- Gráficos de influenciadores identificados, com informações como ID da página, nome da página no Facebook e métricas de influência (por exemplo, número de "likes" recebidos, importância na intermediação de conexões).

Descrição da Solução

Para identificar influenciadores dentro da rede, inicialmente, destacam-se as 10 páginas com o maior número de "likes". Entretanto, essa métrica isolada não é suficiente para uma análise completa dos influenciadores. Em seguida, é feita a identificação de comunidades na rede, a fim de buscar padrões mais profundos de interações.

Além disso, é feito o cálculo da centralidade de intermediação, para destacar páginas que atuam como intermediárias em conexões. Removemos estrategicamente as cinco páginas mais relevantes como pontes entre comunidades. Esse processo de remoção permitiu observar a resiliência da rede, evidenciada pelo aumento no número de componentes conectados. Essa análise revela quais páginas desempenham um papel crucial na conexão entre diferentes comunidades temáticas.

A partir dessas análises combinadas, podemos definir influenciadores de maneira mais abrangente no contexto estudado.

Software Desenvolvido

Linguagem de Programação:

A solução é implementada em Python, uma linguagem de programação de alto nível e amplamente utilizada para tarefas de análise de dados.

Bibliotecas Utilizadas:

Pandas: Utilizada para a manipulação e análise de dados tabulares. Neste caso, para carregar e manipular os dados dos nós e arestas da rede.

Matplotlib: Utilizado para a criação de visualizações gráficas, como o gráfico de barras e o gráfico de resiliência da rede.

Networkx: Utilizado para criar, manipular e estudar a estrutura, dinâmica e funções de redes complexas.

Community: Biblioteca que fornece implementações de algoritmos de detecção de comunidades em redes

Leitura de Dados:

Os dados dos nós e arestas são lidos a partir de arquivos CSV usando a biblioteca pandas. Os nós contêm informações sobre as páginas, enquanto as arestas representam as conexões entre essas páginas.

Análise da Rede:

Utilização do networkx para criar um grafo não direcionado a partir dos dados de arestas. Em seguida, são realizadas análises, como detecção de comunidades usando o algoritmo Louvain. Esse algoritmo é utilizado para detecção de comunidades em redes complexas. O processo envolve a atribuição inicial de cada nó a uma comunidade, seguida por iterações de mesclagem de comunidades para otimizar a modularidade da rede. O algoritmo continua até que a modularidade não possa ser mais aprimorada. A fórmula da modularidade (Q) é geralmente expressa como:

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

Onde:

- A_{ij} é a entrada na matriz de adjacência indicando se há uma aresta entre os nós i e j .
- k_i e k_j são os graus dos nós i e j .
- m é o número total de arestas na rede.
- c_i e c_j são as comunidades atribuídas aos nós i e j .
- $\delta(c_i, c_j)$ é uma função delta que é 1 se $c_i = c_j$ (ou seja, os nós pertencem à mesma comunidade) e 0 caso contrário.

Também é feito o cálculo de centralidade de intermediação. O cálculo da centralidade de intermediação (betweenness centrality) usando a função `nx.betweenness_centrality(G)` na biblioteca NetworkX segue o algoritmo de Brandes para redes. A centralidade de intermediação mede a importância relativa de um nó como intermediário no caminho mais curto entre dois outros nós na rede.

A fórmula para o cálculo da centralidade de intermediação de um nó v é dada por:

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

onde:

- σ_{st} é o número total de caminhos mais curtos de s para t .
- $\sigma_{st}(v)$ é o número de caminhos mais curtos de s para t que passam por v .
- s, t , e v são nós distintos na rede.

Avaliação da Resiliência:

Implementação de uma função (`evaluate_resilience`) para avaliar a resiliência da rede removendo estrategicamente alguns nós e observando o impacto na conectividade da rede.

Visualização Gráfica:

Uso do matplotlib para criar visualizações gráficas, como gráficos de barras e gráficos de resiliência da rede. A função `plt.tight_layout()` é utilizada para garantir um layout adequado nos subplots. A função `plt.savefig()` salva as exibições em um arquivo pdf chamado "graphs.pdf" que é criado dentro da pasta em que o algoritmo é executado.

Instruções de Compilação

O Makefile presente no projeto é destinado a criar e gerenciar um ambiente virtual Python, instalar as dependências especificadas no arquivo `requirements.txt`, e executar o script `analise-comunidades.py`.

Criação do Ambiente Virtual e Instalação de Dependências:

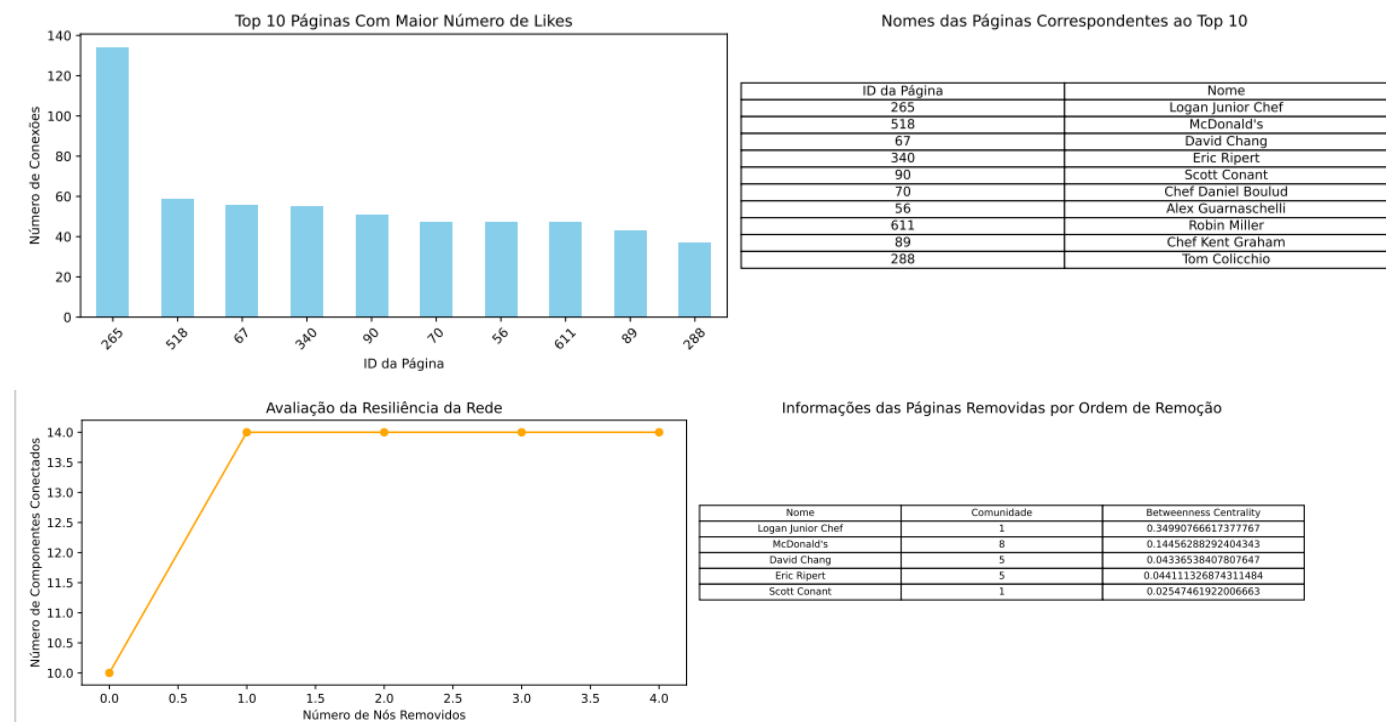
Execute o comando "`make`" para criar o ambiente virtual e instalar as dependências. Isso ativará o ambiente virtual, instalará as dependências do projeto definidas no `requirements.txt` e preparará o ambiente para a execução.

Execução do Projeto:

Após a criação do ambiente virtual, você pode executar o projeto usando o seguinte comando para que seja usado como parâmetro o grafo completo: `make run`

Os arquivos do grafo completo são: "fb-pages-food.nodes" e "fb-pages-food.edges", mas eles estão como default e não é necessário passar como argumento na linha de comando. Para usar algum outro argumento, como por exemplo a amostra do grafo, utilizar: **make run nodes="amostra-vertices.nodes" edges="amostra-arestas.edges"** o arquivo com os vértices vem primeiro, e as arestas depois.

Resultados:



Com base nas análises dos gráficos e tabelas, é evidente que a página "Logan Júnior Chef" exerceu a maior influência no cenário culinário do Facebook em novembro de 2017. Essa conclusão é respaldada por diversos indicadores. Primeiramente, a página apresenta um número expressivamente superior de "likes" em comparação com as demais na rede. Além disso, ao ser estrategicamente removida, observamos um notável aumento no número de componentes conectadas na rede, indicando seu papel crucial como uma ponte entre diferentes comunidades. A relevância de "Logan Júnior Chef" vai além da quantidade de "likes". Logan Guleff, vencedor da segunda temporada do MasterChef Junior americano em 2014, é uma figura multifacetada: personalidade televisiva, autor de livros de receitas e empresário proprietário do Underground Supper Club de Logan. Sua vitória no MasterChef confere uma influência significativa, permitindo o despertar de interesse de empresas do setor para possíveis parcerias. Além disso, sua ampla expertise na área culinária é incontestável.

Em segundo lugar, destaca-se a influência da extensa rede de fast food McDonald's. O McDonald's, como parte integrante da cultura global de fast food, possui uma presença massiva e uma base de fãs considerável. Sua inclusão na análise destaca a importância das grandes marcas no contexto da influência nas redes sociais, especialmente no cenário culinário.

Em terceiro lugar, figura David Chang, chef e fundador do renomado restaurante Momofuku. Com uma série de premiações, incluindo cinco James Beard Awards, Chang é reconhecido como uma das

pessoas mais influentes do século XXI. Sua produção para a Netflix, com séries como "Ugly Delicious" e "Café, Almoço e Jantar", evidencia sua influência e impacto no cenário gastronômico.

Limitações e Considerações Finais

Alguns empecilhos foram encontrados durante o desenvolvimento do projeto, e alguns são válidos destacar. De início, a escolha da base de dados e o escopo a ser tratado foi um processo demorado, visto que nem todas as bases possuem informações suficientes para que seja possível construir as ideias que tive. Porém, após definido, o trabalho fluiu bem, apenas tendo alguns pequenos problemas relacionados ao subsistema windows para linux, WSL, que estive utilizando para desenvolver o trabalho, principalmente relacionado ao Makefile. Ao final, foi de extremo aprendizado, cada dificuldade que tive durante o processo me fez aprofundar ainda mais em conceitos ou tecnologias que não estava acostumada. Também aprendi novos algoritmos, que até então não conhecia e utilizei no trabalho, me empolguei e dediquei realmente durante todo o processo, e estou satisfeita com os resultados obtidos.